Proceedings of PACLIC 19, the 19th Asia-Pacific Conference on Language, Information and Computation.

Analysis of Machine Translation Systems' Errors in Tense, Aspect, and Modality

Masaki Murata
National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gun
Kyoto, 619-0289, Japan
murata@nict.go.jp

Kiyotaka Uchimoto
National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gun
Kyoto, 619-0289, Japan
uchimoto@nict.go.jp

Qing Ma
Ryukoku University
Otsu, Shiga, 520-2194, Japan
qma@math.ryukoku.ac.jp

Toshiyuki Kanamaru
Kyoto University
Yoshida-Nihonmatsu-cho, Sakyo-ku
Kyoto, 606-8501, Japan
kanamaru@hi.h.kyoto-u.ac.jp

Hitoshi Isahara
National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gun
Kyoto, 619-0289, Japan
isahara@nict.go.jp

Abstract

Errors of the translation of tense, aspect, and modality by machine translation systems were analyzed for six translation systems on the market and our new systems for translating tense, aspect, and modality. The results showed that our systems outperformed the other systems. They also showed that the other systems often produced progressive forms rather than the correct present forms. Our systems rarely made this mistake. Translation systems on the market could thus be improved by incorporating the methods used in our systems. Moreover, error analysis of the translation systems on the market identified information that would be useful for improving them.

1. Introduction

Tense, aspect, and modality are difficult to translate appropriately using machines (Shirai et al., 1990; Kume et al., 1990; Dale et al., 2000; Nirenburg et al., 2002). We investigated the error patterns produced by translation systems when translating Japanese tense, aspect, and modality expressions into English. We compared the performance of six translation systems on the market and our new translation systems for tense, aspect, and modality. We found that our systems outperformed the other systems, and we detected error patterns that the other systems often made and our systems rarely made. These results can be used to improve translation systems on the market. Moreover, we extracted error patterns peculiar to each translation system. These errors can be corrected easily because their corresponding sentences can be translated correctly by other systems. These results are useful for improving each translation system.

2. Method

In our investigation, we considered that the translation of tense, aspect, and modality from Japanese to English means the production of the surface expressions of tense, aspect, and modality of the main verb phrase in the English translated sentence. We calculated the accuracy rates and extracted the error patterns in the translations.

We used combinations of the following categories as the surface expressions of tense, aspect, and modality. We refer to the categories as the categories of tense, aspect, and modality.

1. all combinations of {present tense, past tense}, {progressive, non-progressive}, and {perfect, non-perfect} (8 categories)
2. imperative mood (1 category)
3. auxiliary verbs ({present tense, past tense} of “be able to”, {present tense, past tense} of “be going to”, {present tense, past tense} of “be to”, “can”, “could”, and {present tense, past tense} of “have to”, “had better”, “may”, “might”, “must”, “need”, “ought”, “shall”, “should”, “used to”, “will”, and “would”) (21 categories)
4. noun phrases (one category)
5. participial constructions (one category)
6. verb ellipses (one category)
7. interjections or greeting sentences (one category)

We used 800 sentences extracted from a corpus[1] containing 40,198 sentences for the evaluation. We calculated the accuracy rates of six translation systems on the market and our new translation systems and examined the error patterns in the results.

The six translation systems were the latest of leading translation system companies as of October 2003.

Our systems for translating tense, aspect, and modality are based on support vector machines (SVMs) (Murata et al., 2001).[2] They translate Japanese tense, aspect, and modality expressions into English. They detect categories of tense, aspect, and modality previously defined from English expressions. The categories are detected as a categorization problem by SVMs (Cristianini and Shawe-Taylor, 2000; Kudoh, 2000). However, an SVM can handle only two categories at a time. Therefore, we used a pairwise method in addition to the SVM to handle more than two categories (Moreira and Mayoraz, 1998). As training sentences, we used the sentences remaining after eliminating the 800 evaluation sentences from the 40,198-sentence corpus.

We used two feature sets for the machine learning.

- Feature Set 1
  This set consisted of 1- to 10-gram strings at the ends of the input Japanese sentences, e.g., shinai (do not), shinakatta (did not).
- Feature Set 2
  This set consisted of all of the morphemes in each of the input sentences, e.g., *kyou* (today), *watashi* (I), *wa* (topic-marker particle), *hashiru* (run).

---

[1] This corpus was made in our previous studies (Murata et al., 2002b; Murata et al., 2005).

[2] We found that support vector machines were more accurate than other kinds of machine learning methods such as the decision-list method and maximum entropy method (Murata et al., 2001). In addition, the use of support vector machines has been found to be effective in many studies (Taira and Haruno, 2001; Kudo and Matsumoto, 2000; Nakagawa et al., 2001; Murata et al., 2002a). Therefore, we used support vector machines in our translation systems. The detailed parameter settings we used are described in our previous paper (Murata et al., 2001).

Table 1: Occurrence rates of correct categories for tense, aspect, and modality.

Category	Occurrence rate
present	0.65 (516/800)
past	0.45 (356/800)
prefect	0.32 (259/800)
“can”	0.11 (90/800)
“will”	0.11 (87/800)
progressive	0.10 (82/800)
imperative	0.09 (74/800)
“should”	0.07 (59/800)
“must”	0.05 (43/800)
“would”	0.05 (37/800)
past progressive	0.04 (35/800)
perfect progressive	0.04 (28/800)
“ought to”	0.04 (28/800)
“could”	0.03 (23/800)
“may”	0.02 (19/800)
“be going to”	0.02 (18/800)
“had better”	0.02 (13/800)
“shall”	0.01 (12/800)
“have to”	0.01 (11/800)
“be to”	0.01 (10/800)

Table 2: Accuracy rates for translation of tense, aspect, and modality.

Method	Accuracy rate
Baseline	94.50% (756/800)
SVM (all features)	98.75% (790/800)
SVM (Feature Set 1 only)	98.25% (786/800)
SVM (Feature Set 2 only)	94.38% (755/800)
System A	97.00% (776/800)
System B	97.00% (776/800)
System C	95.88% (767/800)
System D	95.50% (764/800)
System E	94.75% (758/800)
System F	94.25% (754/800)

We performed the evaluation using both feature sets, using only Feature Set 1, and using only Feature Set 2.

Because the tense, aspect, and modality expressions of a Japanese sentence can be translated into multiple categories of tense, aspect, and modality in English, we used a strict evaluation procedure. The evaluation was performed by an outside company. We first defined the categories of tense, aspect, and modality of the main verb phrase in the English sentence in an original parallel corpus as the correct category. The original parallel corpus contained example sentences taken from a Japanese-English dictionary (Murata et al., 2002b; Murata et al., 2005). We used as candidate

Table 3: Error patterns.

Pattern		Support vector machine				System on market							Sum
Correct cat.	Incorrect cat. from system	All feat.	FS1 only	FS2 only	Sum	A	B	C	D	E	F	Sum	
present	progressive	1	1	2	4	7	7	9	10	4	8	45	49
present	Past	1	2	19	22	2	2	1	2	2	3	12	34
past	Present	1	1	9	11	2	2	5	5	4	1	19	30
“will”	Present	3	3	2	8	3	3	3	3	4	4	20	28
perfect	Present	1	1	5	7	3	3	3	3	4	1	17	24
perfect	progressive	1	0	1	2	3	3	2	2	2	4	16	18
imperative	Present	2	2	1	5	3	3	0	0	5	2	13	18
present	Perfect	0	0	2	2	0	0	2	2	2	8	14	16
present	imperative	1	3	4	8	1	1	1	1	2	1	7	15
progressive	Past	1	2	2	5	2	2	1	1	2	2	10	15
perfect	Past	1	2	0	3	2	2	1	1	2	2	10	13
“can”	Present	2	2	2	6	1	1	1	1	2	1	7	13
“should”	Present	1	1	1	3	2	2	0	0	4	1	9	12
“would”	Present	1	1	0	2	2	2	1	1	2	2	10	12
past	Perfect	0	0	0	0	0	0	1	2	3	5	11	11
past	past perfect	0	0	0	0	1	1	4	4	1	0	11	11
“must”	Present	1	1	1	3	2	2	0	0	4	0	8	11
“will”	Past	0	0	8	8	0	0	0	0	0	0	0	8
present	“will”	0	0	2	2	0	0	3	2	0	0	5	7
past progressive	past perfect	0	0	0	0	0	0	2	2	3	0	7	7
“can”	Past	0	0	6	6	0	0	0	1	0	0	1	7
past progressive	Perfect	0	0	0	0	0	0	0	0	1	4	5	5
imperative	Past	0	0	4	4	0	0	0	0	0	1	1	5
present	“can”	0	0	0	0	0	0	0	1	0	3	4	4
present	“might”	0	0	0	0	0	0	0	0	3	0	3	3

categories the categories of tense, aspect, and modality in English sentences as translated independently by three professional translators and the categories output by the six translation systems on the market and by our translation systems. Two other professional translators determined whether each candidate category was correct or not. The ones that were judged to be correct were defined as the correct categories. When the two judges disagreed about whether a candidate category was correct or not, it was defined as correct because we examined errors that could be judged to be clearly incorrect. However, we defined as incorrect a candidate category that was judged to be correct only when we assumed a special context or situation.

The occurrence rates for the correct categories are shown in Table 1. The categories for which the frequency was less than ten are not shown. Because more than one category can be correct, the total rates can be more than 1.

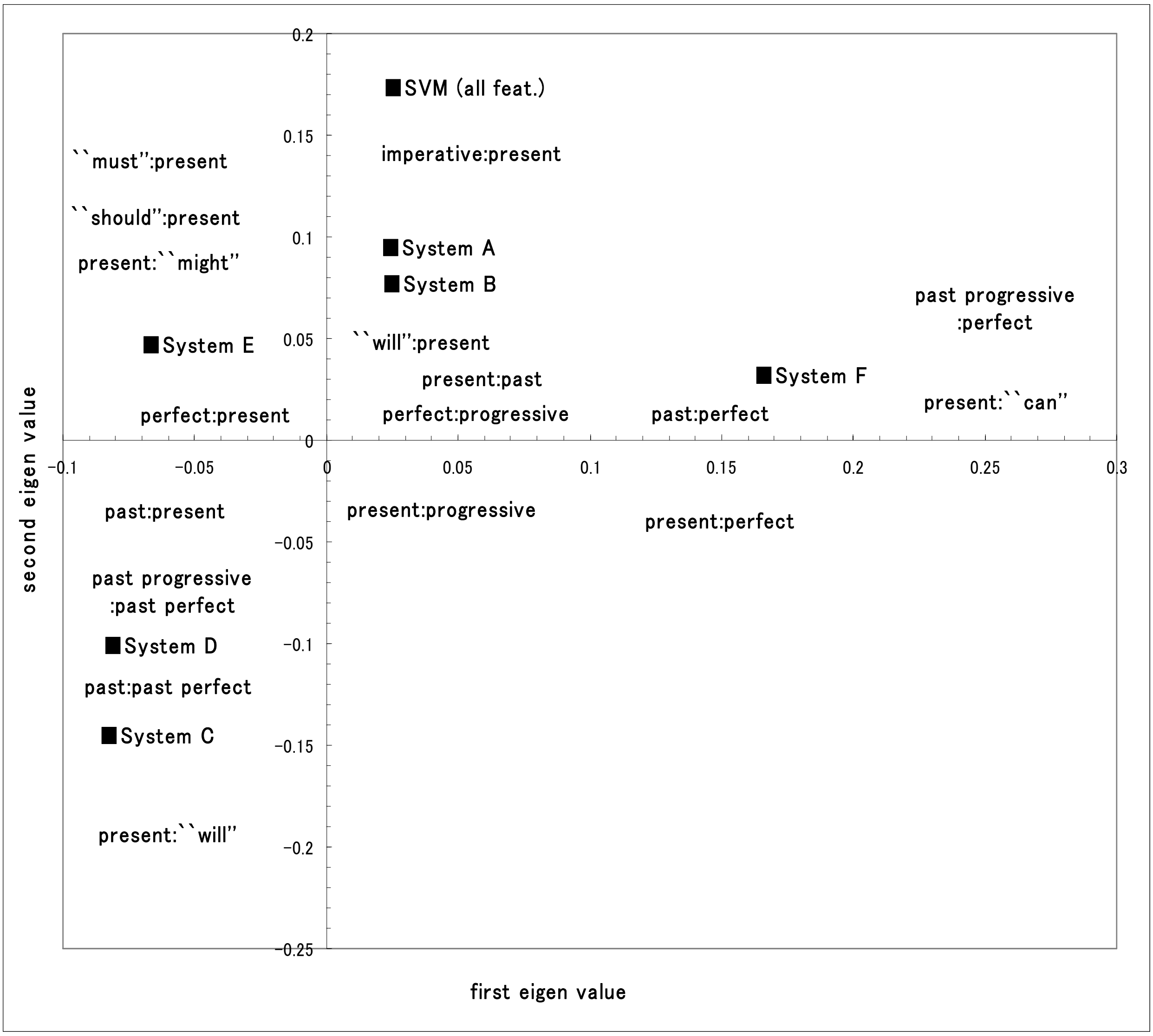

Figure 1: Relationship between translation systems and error patterns.

3. Evaluation and Error Analysis

3.1. Investigation

We evaluated the performance of the translation systems by using the method described in the previous section. The accuracy rates are shown in Table 2. For the baseline method, if a sentence ended with *ta* (a Japanese particle used for the past tense), it was judged to be in the past tense; otherwise, it was judged to be in the present tense. When a translation system could not output a sentence, the output of the baseline method was used instead. We refer to the six translation systems as A, B, C, D, E, and F.

As shown in Table 2, the SVM had the highest accuracy rates when all features were used. Systems A and B had the highest accuracy rates of the systems on the market. Systems E and F had accuracy rates near that of the baseline method.

Next, we analyzed errors by investigating the error patterns of the cases where the translations were judged to be incorrect. An error pattern was a pair of the correct category and the incorrect category output by a system. When multiple categories were correct, each case was considered as

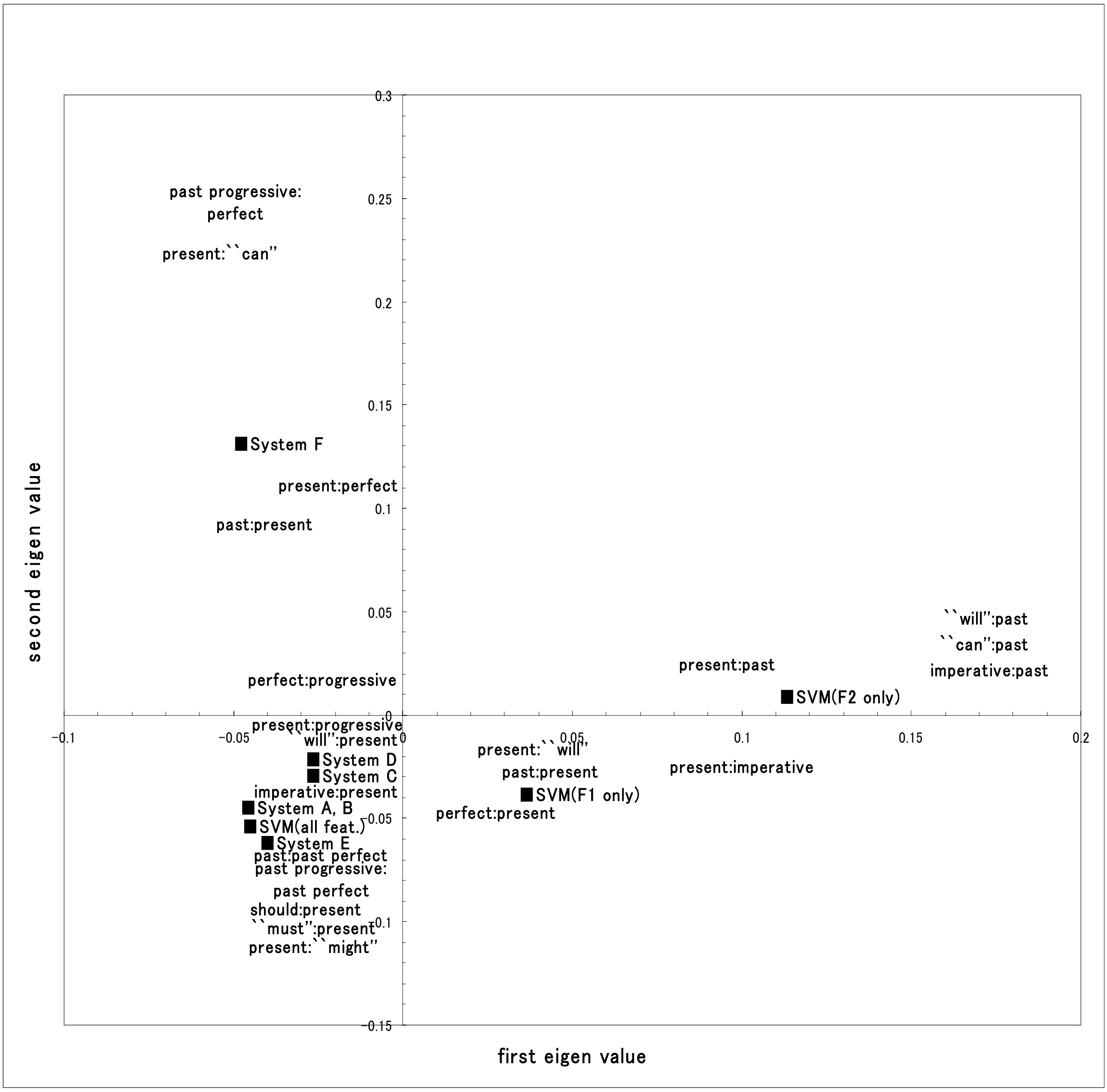

Figure 2: Relationship between translation systems and error patterns (when all three SVMs were used).

an error pattern (e.g., when both “present” and “progressive” were correct, and the system output was “past”, two error patterns, the pair “present” and “past” and the pair “progressive” and “past” were extracted as error patterns.) The category of "no output" was defined for the case when a translation system on the market did not output a verb phrase in the English translation; however, this rarely occurred, so the category is not presented in the tables shown here. The results of investigating the error patterns are shown in Table 3. Only shown are those patterns with a total frequency of more than nine or an error frequency for an individual system of more than two.

We investigated the tendency of the distribution of error patterns for the six translation systems on the market and the SVM when all features were used. (We also used the SVM with Feature Set 1 only and with Feature Set 2 only, but displaying all the results on one graph made the graph and the analysis shown in Figure 2 too complicated for clear understanding.) We extracted those error patterns for which the frequency of errors for a system was more than two and calculated the co-occurrence frequency of the error patterns and the seven translation systems. We constructed cross tables in this manner and then used the dual scaling method to analyze them (Weller and Romney, 1990; Ueda et al., 2003). As shown in Figure 1, the figure also roughly shows the error patterns of each translation system. For example, the proximity of error patterns “past:perfect”, “present:perfect”, and “past progressive:perfect” near System F indicate that System F produced perfect forms rather than the correct past, present, or past progressive forms more often than the other systems.

3.2. Examination of our systems

We first examined the performance of the SVM when only Feature Set 1 was used. It made a few more errors in many error patterns than when both feature sets were used. When only Feature Set 2 was used, many errors were made with the pair “present” and “past”. We found that translating Japanese tense, aspect, and modality expressions is difficult when only word information is used. The characters at the ends of Japanese sentences are also very important.

3.3. Examination of systems on the market

We next examined the performance of the systems on the market. As shown in Table 3, Systems A and B had virtually the same performance. The output categories of tense, aspect, and modality for System A were exactly the same as those for System B, and the error patterns for System A were also exactly the same as those for System B. Although Systems A and B were developed by different companies, because their outputs were very similar, they likely were developed cooperatively. We also found that the translated sentences and the output categories of tense, aspect, and modality were very similar for Systems C and D. Again, some cooperative development appears to have been done. We can predict these relationships by using Figure 1 because Systems A and B are near each other and Systems C and D are near each other.

Systems A and B had the highest accuracy rates, and Systems C and D had the next highest accuracy rates among the six systems on the market. This indicates that cooperative development tends to result in higher accuracy rates.

3.4. Comparison of systems on the market and our systems

We found that the error patterns of the systems on the market very often produced progressive forms rather than the correct present forms, while our systems rarely made such errors. An example of this is as follows.

Input Japanese sentence:

kono	*heya niwa*	*suidou ga*	*torituke rareteiru.*
(this)	(room)	(city water)	(is laid)

Translation result: A water service is being installed on this room.
Correct translation: City water is laid on in this room.

The system produced the progressive form rather than the correct present form. Our systems made this error much less often than the translation systems on the market. We found that the methods used in our systems could alleviate this problem. Use of these methods will thus aid in the development of future machine translation systems.

We also found that our systems made fewer errors in producing a present form rather than the correct past form and in producing a present or progressive form rather than the correct perfect form than the systems on the market. Although the perfect form is thought to be difficult to handle in a translation system, the SVM made few such errors when all features were used. Our system is useful for reducing such errors.

All of the systems on the market and our systems often produced a present form rather than “will”. In Japanese, the same base form is used for both the future and the present. Therefore, correct translation to “will” and the present form is difficult, causing trouble for any Japanese-to-English translation system.

3.5. Examination of each system on the market

Next, we examined the error patterns of each translation system by using Table 3 and Figure 1. We found that Systems C and D more often produced a past perfect form rather than the correct past progressive or past form. An example of this is as follows:

Input Japanese sentence:

kaze de	*ho ga*	*hira hira to*	*hirameite ita*
(wind)	(sail)	(flutteringly)	(was flapping)

Translation result: The sail had flashed flutteringly by the wind.
Correct translation 1: The wind was flapping the sails.
Correct translation 2: The flag fluttered in the wind.

The systems produced the past perfect form rather than the correct past progressive or past form. This is a typical error with Systems C and D. Apparently, the systems failed to adjust, so they were more likely to produce a past perfect form. Engineers constructing such systems should be able to improve their performance relatively easily by examining these results. In addition, Systems C and D often produced “will” rather than the correct present form. An example of this is as follows.

Input Japanese sentence:

sadou no	*kigen wa*	*16 seiki*	*izen ni made*	*saka noboru.*
(tea ceremony)	(origin)	(16th century)	(before)	(trace)

Translation result: The origin of the tea ceremony will go back even before the 16th century.
Correct translation 1: The origin of the tea ceremony dates back before the 16th century.
Correct translation 2: The origin of the tea ceremony can be traced back before the 16th century.
Correct translation 3: The tea ceremony originated before the 16th century.

Alleviating this problem, however, is difficult because of the translation problem with respect to “will” and present forms as described above.[3]

System E more often produced “might” rather than the correct present form. An example of this is as follows.

Input Japanese sentence:

sate,	*soudan*	*shitai koto ga*	*aru noda.*
(well)	(consult)	(want)	(I have)

[3] This example shows another problem of the translation of tense, aspect, and modality. In this example, the categories of tense, aspect, and modality in the correct answers correlate with the verbs: “Can” is only good with “traced back”, not “originate” or “date”. A past tense is only good with “originate”, not “can be traced” or “date”. A present tense is only good with “date”, not “originate”. In this study, we only handled the categories of tense, aspect, and modality, but we should also handle verbs correlated with these expressions. In a future study, we would like to handle the categories of tense, aspect, and modality and verbs correlated with them.

Translation result: Well, I might want to consult.
Correct translation 1: Now, I have something to talk over with you.
Correct translation 2: May I ask your advice?

In this context, the use of “might” is unnatural. In addition, System E more often produced a present tense rather than the correct “must” or “should”. An example of this is as follows.

Input Japanese sentence:

kimi no	*souiu*	*okonai wa*	*togame rareru*	*bekida.*
(your)	(such)	(conduct)	(be blame)	(must be)

Translation result: It is necessary to blame such your doing.
Correct translation 1: You must be blamed for such conduct.
Correct translation 2: You should be blamed for such conduct.

Although System E produced a meaning similar to the correct modality by using the sentence pattern “it is necessary”, “should” or “must” would be more appropriate expressions. The developers of System E should be able to improve the performance of their system relatively easily by examining these results.

System F more often produced a perfect form rather than the correct present, past, or past progressive form. An example of this is as follows.

Input Japanese sentence:

sore wa	*watashi no*	*konomi ni*	*atte iru.*
(it)	(my)	(taste)	(suit)

Translation result: It has matched my taste.
Correct translation: It suits my taste.

A present form is appropriate. Apparently, the system failed to adjust, so it was more likely to produce a perfect form. In addition, when we examined the translation results for System F, we found cases where the translation was incorrectly divided into two sentences. An example of this is as follows.

Input Japanese sentence:

kare wa	*jissai yori*	*wakaku*	*mieru.*
(he)	(he really is)	(younger)	(look)

Translation result: It is younger than practice and he can be seen.
Correct translation: He looks younger than he really is.

The sentence was incorrectly divided into two sentences, and the system produced “can” rather than the correct present form. The developers of System F should easily be able to improve the performance of their system by examining these results.

3.6. Other kinds of error patterns

We also examined other error patterns. The SVM and Systems A, B, E, and F sometimes produced a present form rather than the correct imperative form, while Systems C and D did not. Systems C and D should thus produce better translations in such cases. An example of this is as follows.

Input Japanese sentence:

Seikou	*shitai*	*nara,*	*ima noyouni*	*damin wo*	*musabotte*	*itewa dameda.*
(succeed)	(want)	(if)	(now)	(idleness)	(live in)	(don't/stop)

Translation result: When the idle slumber is coveted as today if it wants to succeed, it is useless.
Correct translation: If you are to succeed, stop idling away your time, and work hard.

The SVM and Systems A, B, E, and F apparently had trouble translating *itewa dame* (don't/stop). The developers of these systems should be able to easily improve their performance by examining these results.

Systems A, B, C, D, and E often produced a present form rather than the correct past or perfect form. An example of this is as follows.

Input Japanese sentence:

saibankan ga	*nyuutei*	*shita.*
(judge)	(enter a court room)	(did)

Translation result: The enter the court room of the judge.
Correct translation: The judges entered the court room.

In the translation result, the subject noun phrase is missing. Because the structure of the translated sentence was broken, the tense, aspect, and modality expressions were broken. This error pattern would be difficult to eliminate. Doing so would require improving the overall performance of the translation system.

4. Conclusion

We analyzed errors in the Japanese-to-English translation of tense, aspect, and modality by six machine translation systems on the market and our new translation systems for tense, aspect, and modality.

Our evaluation showed that our support vector machine (SVM) using all features had the highest accuracy rate. Two of the systems on the market had the highest accuracy rates among those on the market, while two others had accuracies as low as that of the baseline method.

Error analysis showed that when the string characters at the ends of sentences were not used, the SVM had low accuracy and often produced a past form rather than the correct present form. Use of the string characters at the ends of sentences is thus important.

Our system outperformed the other systems. The other systems often produced progressive forms rather than the correct present forms, while our system rarely made such errors. This indicates that the translation systems on the market can be improved by using the methods used in our system. We also found that our systems made fewer errors in producing a present form rather than the correct past form and in producing a present or progressive form rather than the correct perfect form. These errors can be corrected by using the methods used in our system.

In our experimental results, all of the systems on the market and our systems often produced a present form rather than “will”. This indicated that correct translation to “will” and present forms is difficult. This paper is thus useful for identifying error patterns that are difficult to correct in translation systems.

Our investigation detected error patterns made by each system on the market. Most of these errors can be corrected relatively easily because their corresponding sentences can be translated correctly by the other systems. This paper is thus useful for identifying error patterns that can be relatively easily corrected in each translation system.

We compared the results of individual systems. We extracted error patterns that are difficult to correct by extracting error patterns that were made in almost all of systems. We also extracted error patterns that are relatively easy to correct by extracting error patterns that were made in a few systems and not made in other systems. Our approach is useful for identifying whether each error pattern can be corrected easily.

5. Acknowledgements

We are grateful to the machine translation systems engineers who made valuable comments about this paper.

6. References

Nello Cristianini and John Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Cambridge University Press.

Robert Dale, Hermann Moisl, and Harold Somers. 2000. *Handbook of Natural Language Processing.* Marcel Dekker, Inc.

Taku Kudo and Yuji Matsumoto. 2000. Japanese dependency analysis based on support vector machines. *EMNLP/VLC 2000.*

Taku Kudoh. 2000. TinySVM: Support Vector Machines. http://cl.aist-nara.ac.jp/taku-ku/software/TinySVM/index.html.

Masako Kume, Takayuki Toyoshima, and Masaaki Nagata. 1990. Japanese aspect processing for spoken language translation. In *Information Processing Society of Japan, the 40th National Convention, 1F-7*, pages 415-416. (in Japanese).

Miguel Moreira and Eddy Mayoraz. 1998. Improved pairwise coupling classification with correcting classifiers. In *European Conference on Machine Learning*, pages 160-171.

Masaki Murata, Kiyotaka Uchimoto, Qing Ma, and Hitoshi Isahara. 2001. Using a support-vector machine for Japanese-to-English translation of tense, aspect, and modality. *ACL Workshop on the Data-Driven Machine Translation.*

Masaki Murata, Qing Ma, and Hitoshi Isahara. 2002a. Comparison of three machine-learning methods for Thai part-of-speech tagging. *ACM Transactions on Asian Language Information Processing*, 1(2):145-158.

Masaki Murata, Masao Utiyama, Kiyotaka Uchimoto, Qing Ma, and Hitoshi Isahara. 2002b. Correction of errors in a modality corpus used for machine translation by using machine-learning method. In *TMI 2002*.

Masaki Murata, Masao Utiyama, Kiyotaka Uchimoto, Qing Ma, and Hitoshi Isahara. 2005. Correction of errors in a verb modality corpus used for machine translation with a machine-learning method. *ACM Transactions on Asian Language Information Processing.* (to appear).

Tetsuji Nakagawa, Taku Kudoh, and Yuji Matsumoto. 2001. Unknown word guessing and part-of-speech tagging using support vector machine. In *NLPRS 2001*.

Sergei Nirenburg, Harold Somers, and Yorick Wilks. 2002. *Readings in Machine Translation.* Bradford Books.

Satoshi Shirai, Akio Yokoo, and Francis Bond. 1990. Generation of tense in newspaper translation. In *The Institute of Electronics, Information and Communication Engineers, Autumn Convention*, page D-69. (in Japanese).

Hirotoshi Taira and Masahiko Haruno. 2001. Feature selection in svm text categorization. In *Proceedings of AAAI2001*, pages 480-486.

Taichiro Ueda, Masao Karita, and Kazue Honda. 2003. *Jissen Workshop Excel Tettei Katsuyou Tahenryou Kaiseki*. Shuuwa System. (in Japanese).

Susan C. Weller and A. Kimball Romney. 1990. Metric Scaling: *Correspondence Analysis (Quantitative Applications in the Social Sciences).* SAGE Publications.